Cluster-based characterization of multi-dimensional tropospheric ozone variability in coastal regions: an analysis of lidar measurements and model results

Claudia Bernier[1], Y. Wang[1], G. Gronoff[2], T. Berkoff[2], E. Knowland[4,5], J. T. Sullivan[4], R. Delgado[6], V. Caicedo[6], B. Carroll[6]

[1] Department of Earth and Atmospheric Science, University of Houston, Houston, Texas, US, [2] NASA Langley Research Center, Hampton, VA, 23666, USA, [3] Science Systems and Application Inc., Hampton, VA, 23666, USA, [4] NASA Goddard Space Flight Center, Global Modeling and Assimilation Office, Greenbelt, MD, 20771, USA, [5] Morgan State University, Goddard Earth Science Technology & Research (GESTAR) II, Baltimore, Maryland, USA, [6] UMBC

AMS 2022 Oral Presentation

Session: Addressing Air Quality Challenges at Urban–Land–Water Interfaces during Recent Field Studies. Part I

UNIVERSITY of HOUSTON

EARTH AND ATMOSPHERIC SCIENCES

Goal: Use multi-dimensional lidar measurements from campaigns to evaluate simulated coastal O_3 events

-- Directly address difficulties of measuring, modeling, & forecasting the large variability of O_3 in coastal environments

-- Use 2-D high temporal & fine resolution lidar measurements to help elucidate gaps within current air quality models

-- Complete observing system (surface land & water observations, model simulations)

=> a formidable tool for evaluating air pollution in coastal regions

-- Evaluate the ability of hourly temporal O_3 measurements & improve the abilities to simulate coastal O_3 concentrations

NASA Air quality campaigns

- 3 recent air quality campaigns:
  - OWLETS 1 (O_3 Water-Land Environmental Transition Study): July 5 - August 3, 2017
  - OWLETS 2: June 6 - July 6, 2018
  - LISTOS (Long Island Sound Tropospheric O_3 Study): July 12 - August 29, 2018

- Suite of detailed airborne & ground measurements measurements (e.g. ozonesonde, aerosol profiling - ground-based ceilometers, Pandora spectrometer, POM, etc.)

- Two <u>TOLNet lidars</u> used:
  - NASA Goddard Space Flight Center (GSFC) Tropospheric Ozone (TROPOZ) Differential Absorption Lidar (DIAL; Sullivan et al. 2014, 2015a)
  - NASA LaRC Mobile Ozone Lidar (LMOL; De Young et al. 2017; Farris et al. 2018)

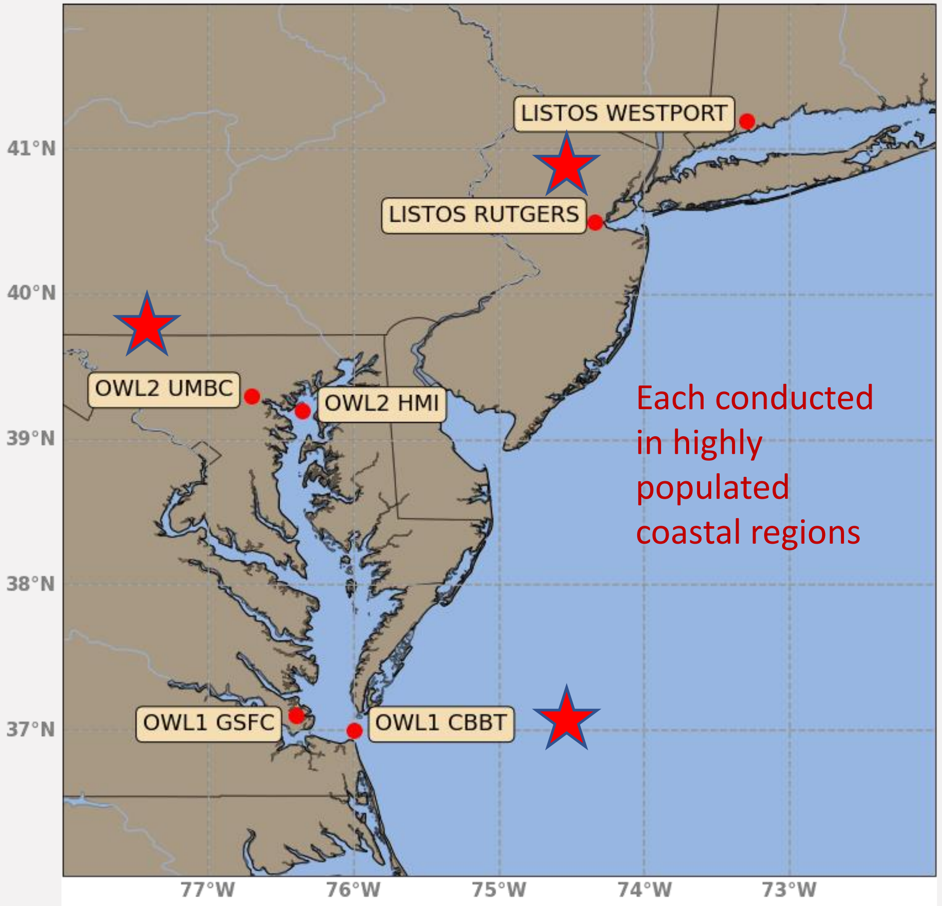
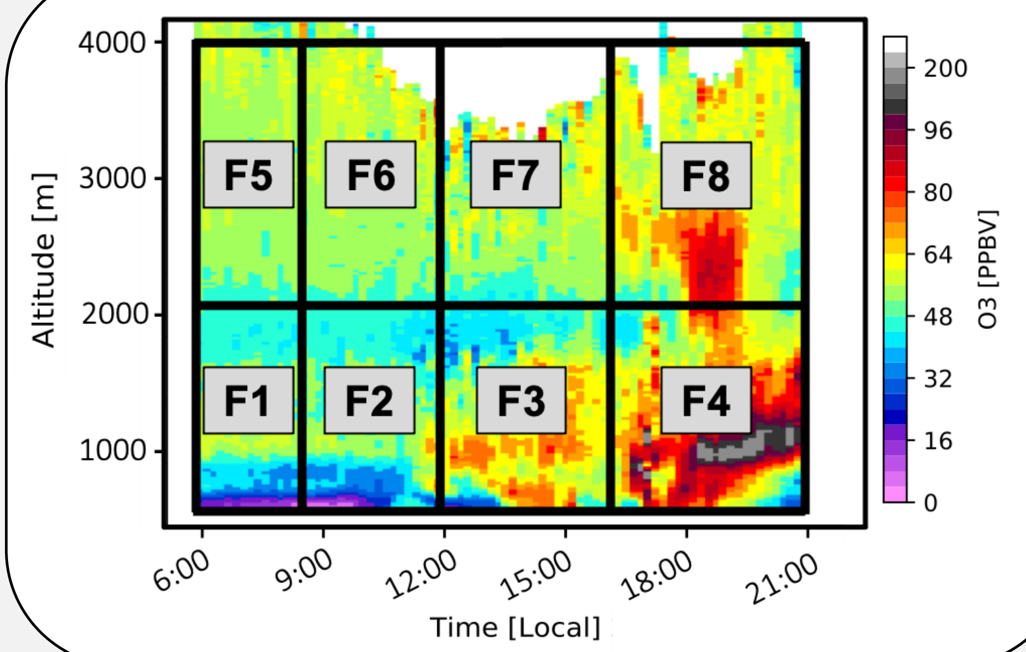


Figure 3. Inset map of Chesapeake Bay airshed in Maryland, VA and Long Island Sound in NY with 6 lidar monitoring locations used for OWLETS-1, OWLETS-2, & LISTOS (highlighted & labeled).

91 individual 2-D profile curtains used from both lidars:
- **26** from OWLETS-1,
- **28** from OWLETS-2,
- **37** from LISTOS
- https://www-air.larc.nasa.gov/index.html

Characterizing O_3 profile curtains



- Color coding *(left)* shows a typical O_3 profile taken from a lidar instrument
  - *August 6, 2018; LMOL at Westport, CT; LISTOS Campaign*
- Developed clustering method for characterizing 2-D O_3 curtain profiles from lidar measurements
- Constrained by the goal of evaluating lower level tropospheric O_3 limitations of the lidar measurements

- Split profiles (91 total profiles) into "slabs" based on **altitude** & **time**
  - F1 – F8 *(above)* indicate input feature ranges
- Calculated average features implemented into K-Means Clustering algorithm

<u>Features:</u>

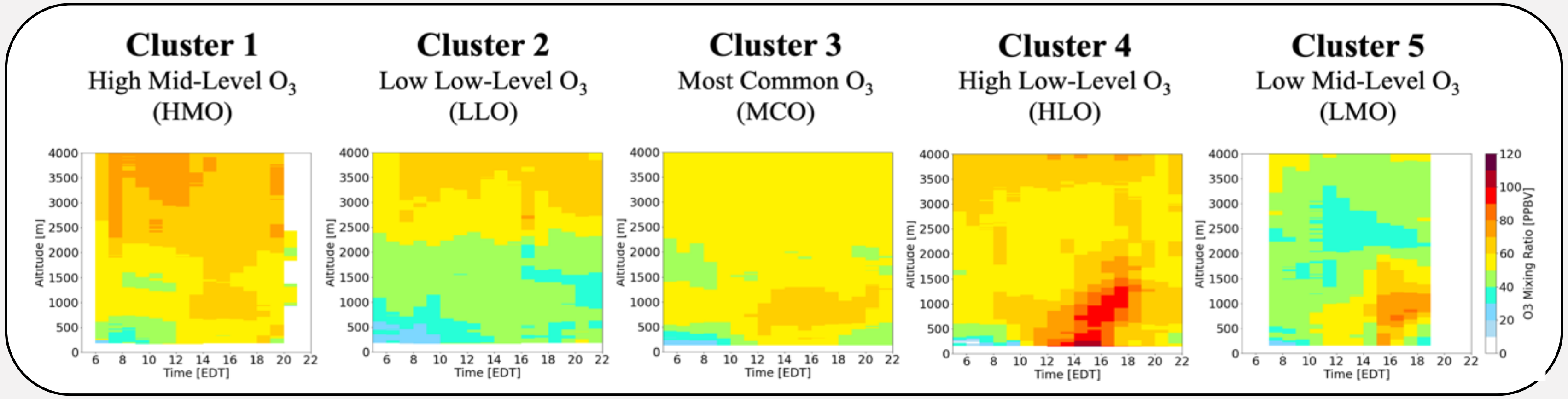
▻ **Altitude 2 sectors:**	▻ **Time 4 sectors:**
• 0 – 2000 m (low-level)	• 6:00 – 8:00
	• 8:00 – 12:00
• 2000 – 4000 m (mid-level)	• 12:00 – 16:00
	• 16:00 – 21:00

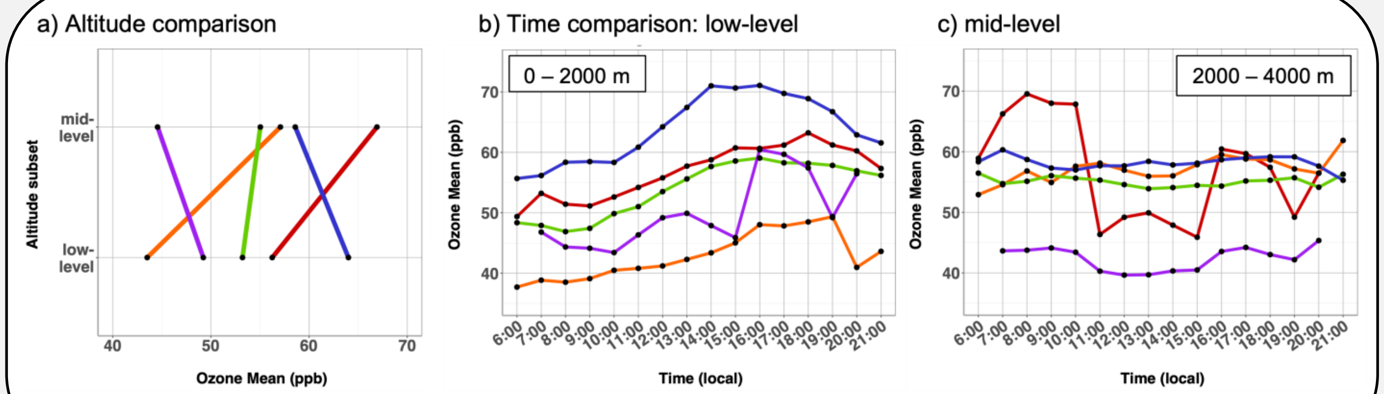
Clustering O_3 profiles

Cluster #	a) No. of profiles	b) O_3 Max (ppb)	c) O_3 Min (ppb)	d) T (°F)	e) WS (m s^{-1})
1	25	86.50	42.17	73.80	1.51
2	14	72.77	28.85	71.42	1.62
3	27	86.64	34.18	77.34	1.30
4	18	97.77	44.08	77.79	1.24
5	5	67.70	29.07	74.93	1.51

Fig. Cluster statistics

- Mean lidar O_3 vertical profiles for each of the clusters



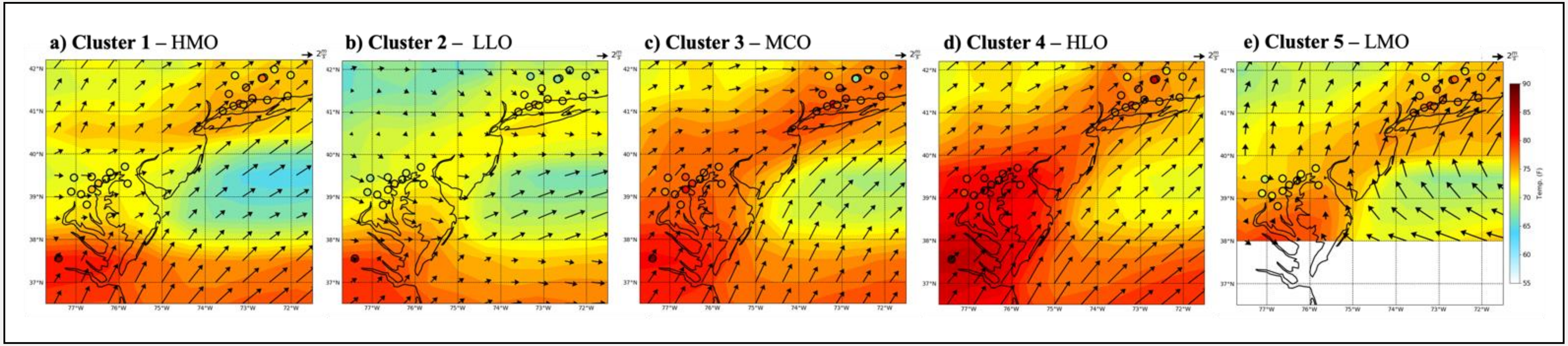


– C1 – C2 – C3 – C4 – C5

Descriptions of clusters:

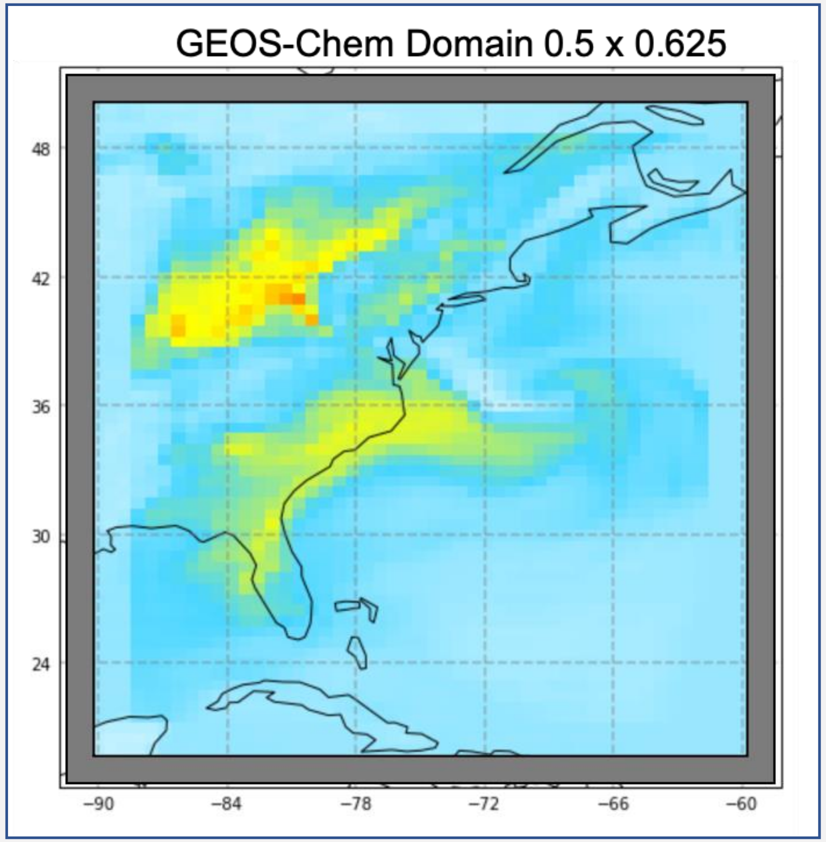
- **Cluster 1** = highest mid-level O_3 (HMO) cluster
- **Cluster 2** = lowest low-level O_3 (LLO) cluster
- **Cluster 3** = most common O_3 (MCO) cluster
- **Cluster 4** = highest low-level O_3 (HLO)
- **Cluster 5** = least common & lowest mid-level O_3 (LMO) cluster

Meteorological surface analyses



- Meteorological conditions support clustered lidar O_3 profiles.
- Cluster 3 & 4 presented higher overall observed & simulated surface temperatures
- Clusters 1, 3, and 4 (highest surface O_3) -- predominant offshore, westerly wind
  - conducive to higher surface O_3 concentrations
- Cluster 2 = lowest temperatures -- cluster with lowest surface O_3 concentrations
- Cluster 5 temperatures moderately high but, in contrast, average wind speed is higher (specific over LIS)
  - wind direction is predominantly onshore (Easterly – Southerly) -- transport of cleaner marine air corroborates lower surface O_3 levels

Model setup



- GEOS-Chem (MERRA2 meteorology) nested 0.5°x 0.625° horizontal resolution ~ using 2°x 2.5° global boundary conditions.
- GEOS Composition Forecasting (GEOS-CF) system 0.25° horizontal resolution data *(https://gmao.gsfc.nasa.gov) (Knowland et al., 2019)*
- Similar chemistry schemes

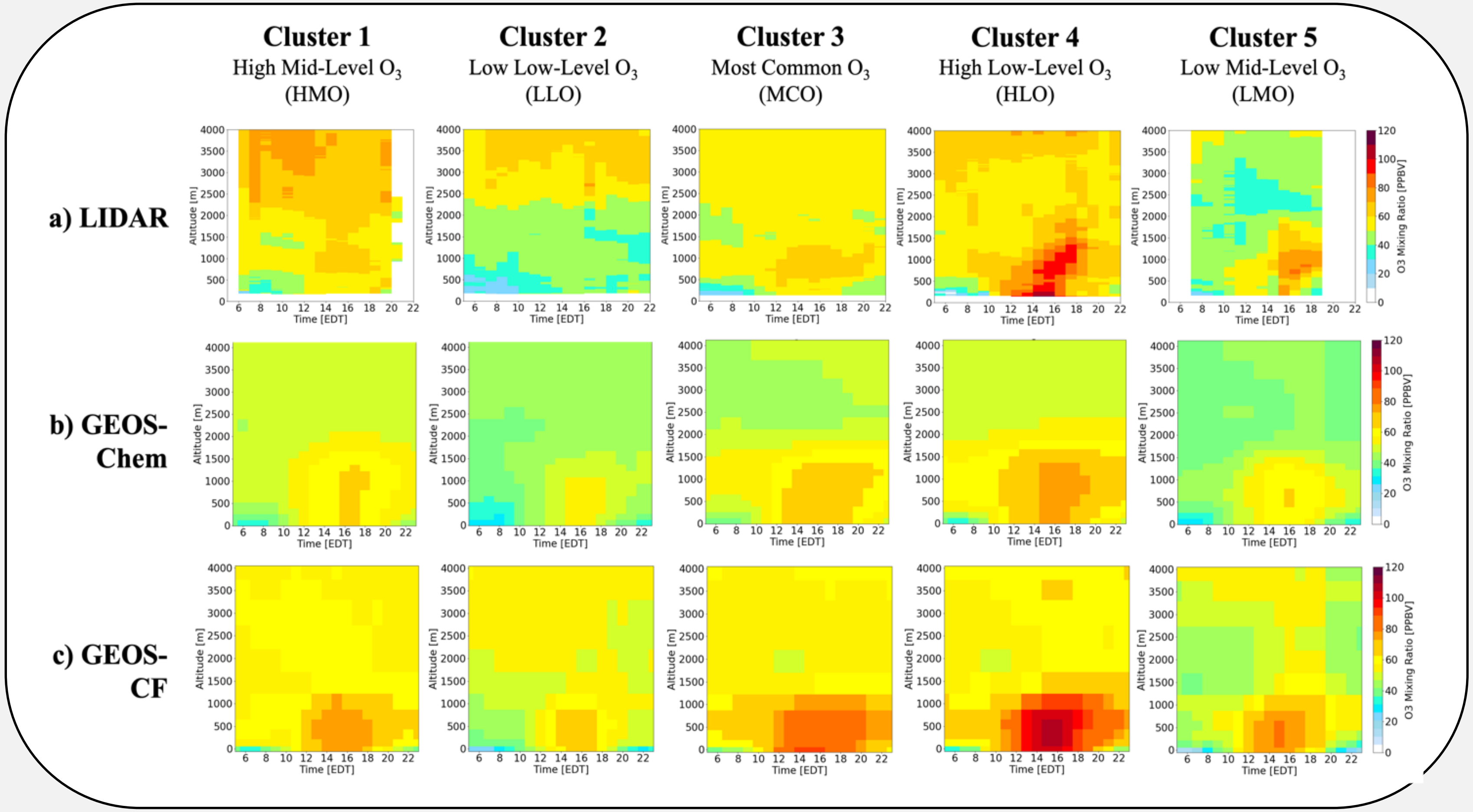
- The main model differences:

1. GEOS-Chem => offline CTM using archived meteorology; GEOS-CF => online, simulates atmospheric composition simultaneously with meteorology
2. GEOS-CF spatial resolution => 0.25°; GEOS-Chem spatial resolution => 0.5° x 0.625°
3. GEOS-CF anthropogenic emissions => Harmonized Gridded Air Pollution (HTAP; v2.2) from the Emission Database for Global Atmospheric Research (EDGAR); GEOS-Chem anthropogenic emissions => CEDS
4. Data only available for <u>2018</u> ~ analysis adjusted for this comparison

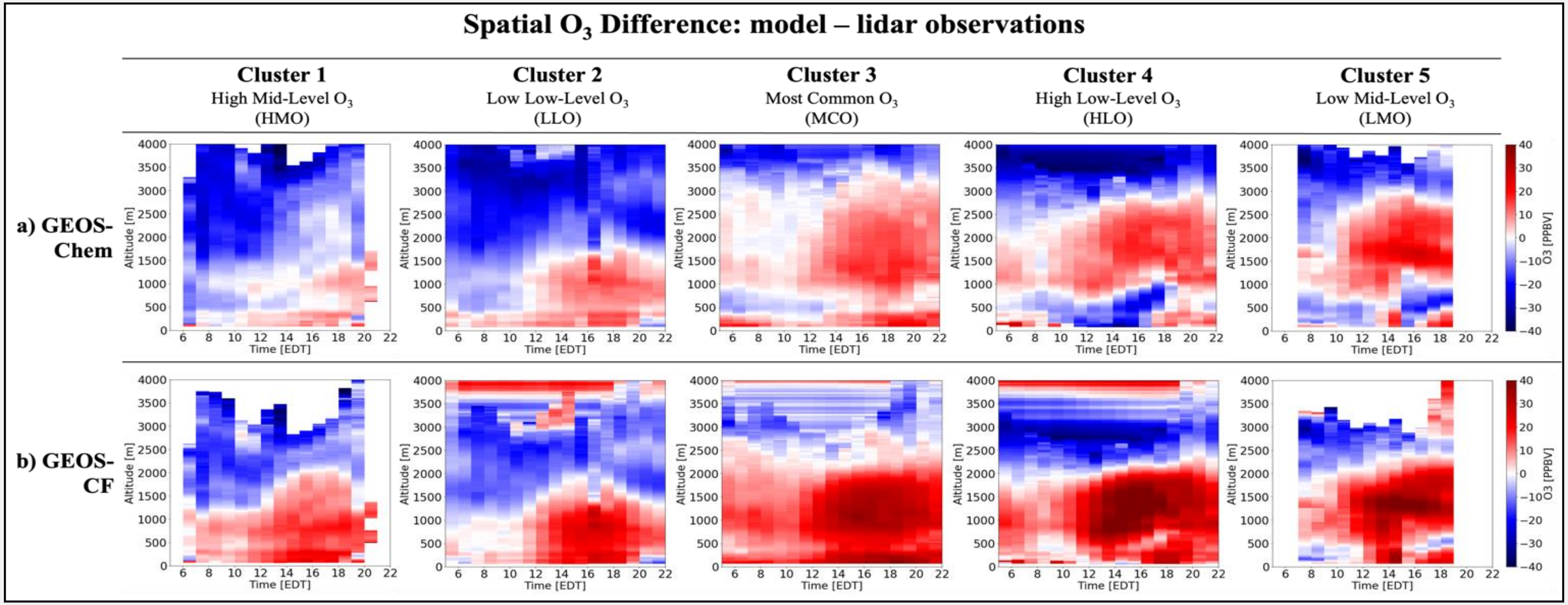
Air Quality Model Results

- Both models simulate low-level O_3 well
- Both models struggle to simulate the magnitude of O_3 in mid-level

- GEOS-Chem simulates low O_3 concentrations best in the low-level altitude (e.g., C2)
- Both models struggled to capture moderately high O_3 in low-level altitudes (e.g., C3)
- Both models underestimate mid-level O_3 with worst performance in C1
- Lower O_3 concentrations are better simulated by both models in mid-level altitudes



Cluster Bias



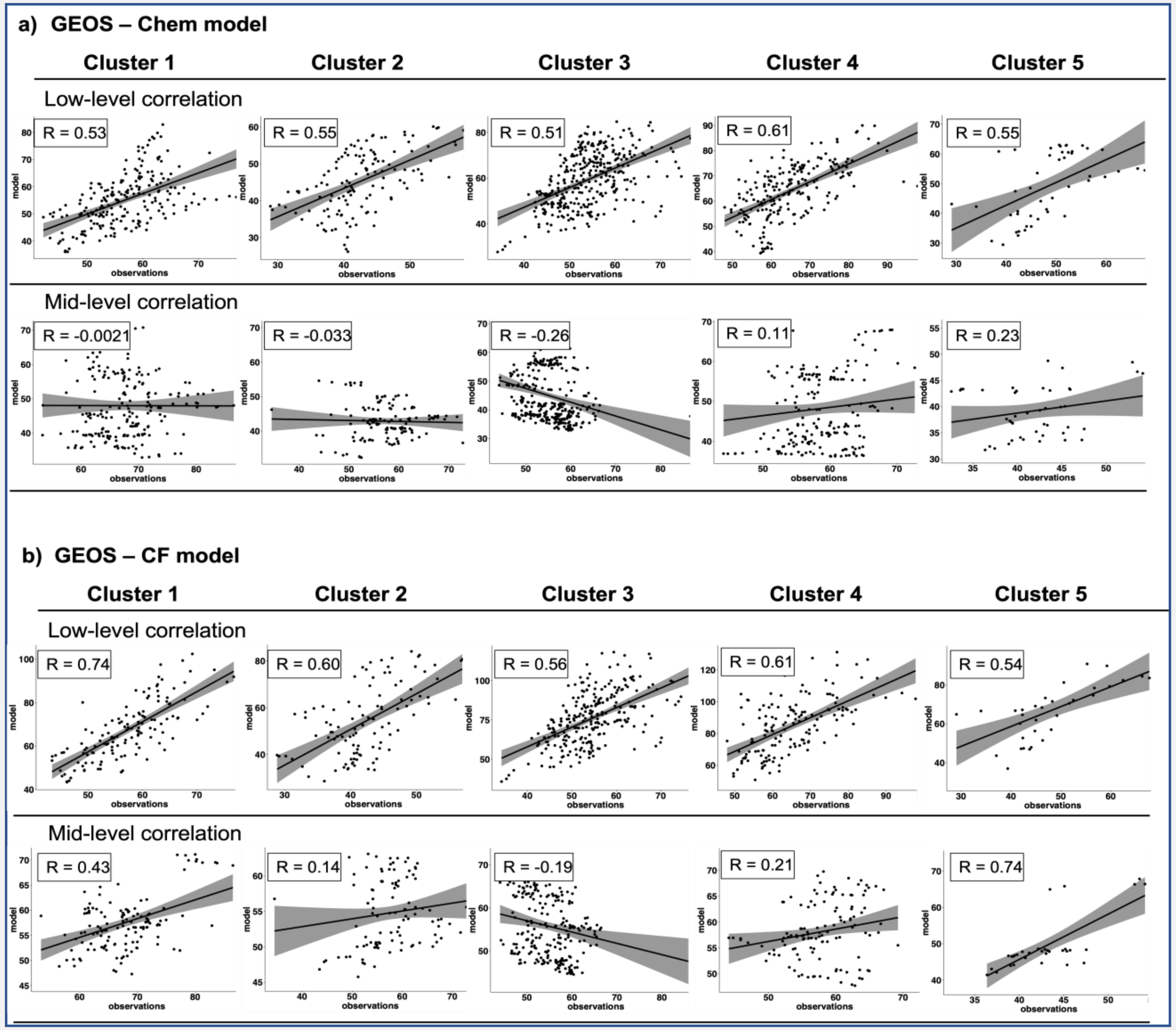
Low-level bias:

- ➢ GEOS-Chem low unsystematic bias range (model – lidar observations)
- ➢ GEOS-CF systematic high **+ bias** but fair relationship

Mid-level bias

- ➢ GEOS-Chem systematic high negative **– bias**
- ➢ GEOS-CF low unsystematic bias

Cluster Bias & Correlation



Bias varies by cluster for both models.

Low-level bias:

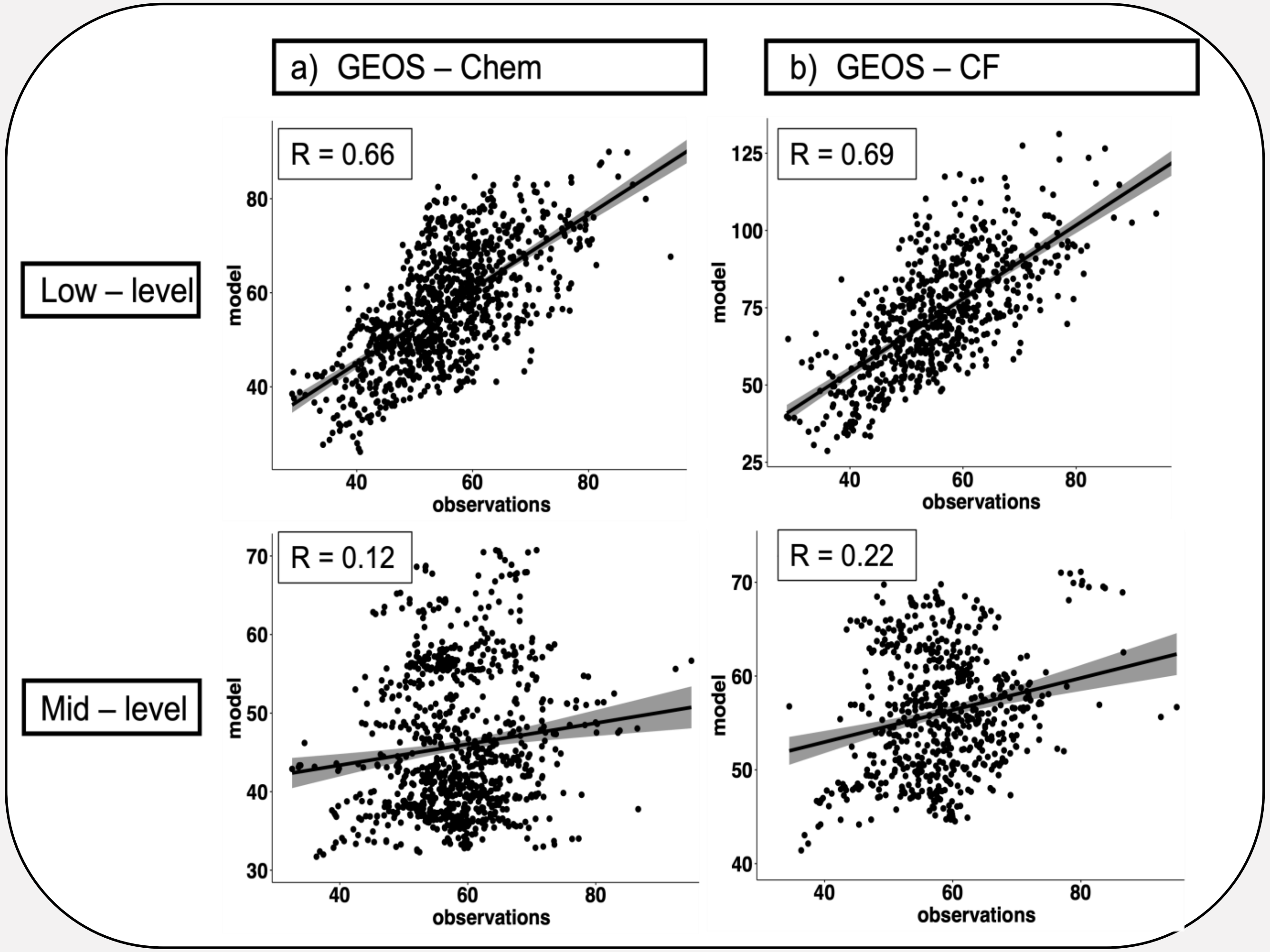
- GEOS-Chem fair relationship with the lidar observations
- Both models able to simulate low-level O_3 pattern well, but GEOS-CF was not able to simulate magnitude consistently overestimating O_3 in low-level altitudes

Mid-level bias:

- GEOS-Chem and GEOS-CF both have a weak relationship with the lidar observations
- Neither simulated mid-level O_3 pattern well, but GEOS-CF was able to simulate magnitude slightly better

Overall Model Correlation

- Overall correlation results indicate GEOS-CF (better grid resolution & an online model) had a slightly better relationship with lidar observations than GEOS-Chem
- There are still limitations to both models especially when simulating mid-level O_3
- Known model errors & coarse horizontal and vertical grid resolution contribute to the difficulty in simulating fine-scale coastal O_3 variability



Impact of meteorological factors on clusters & model performance

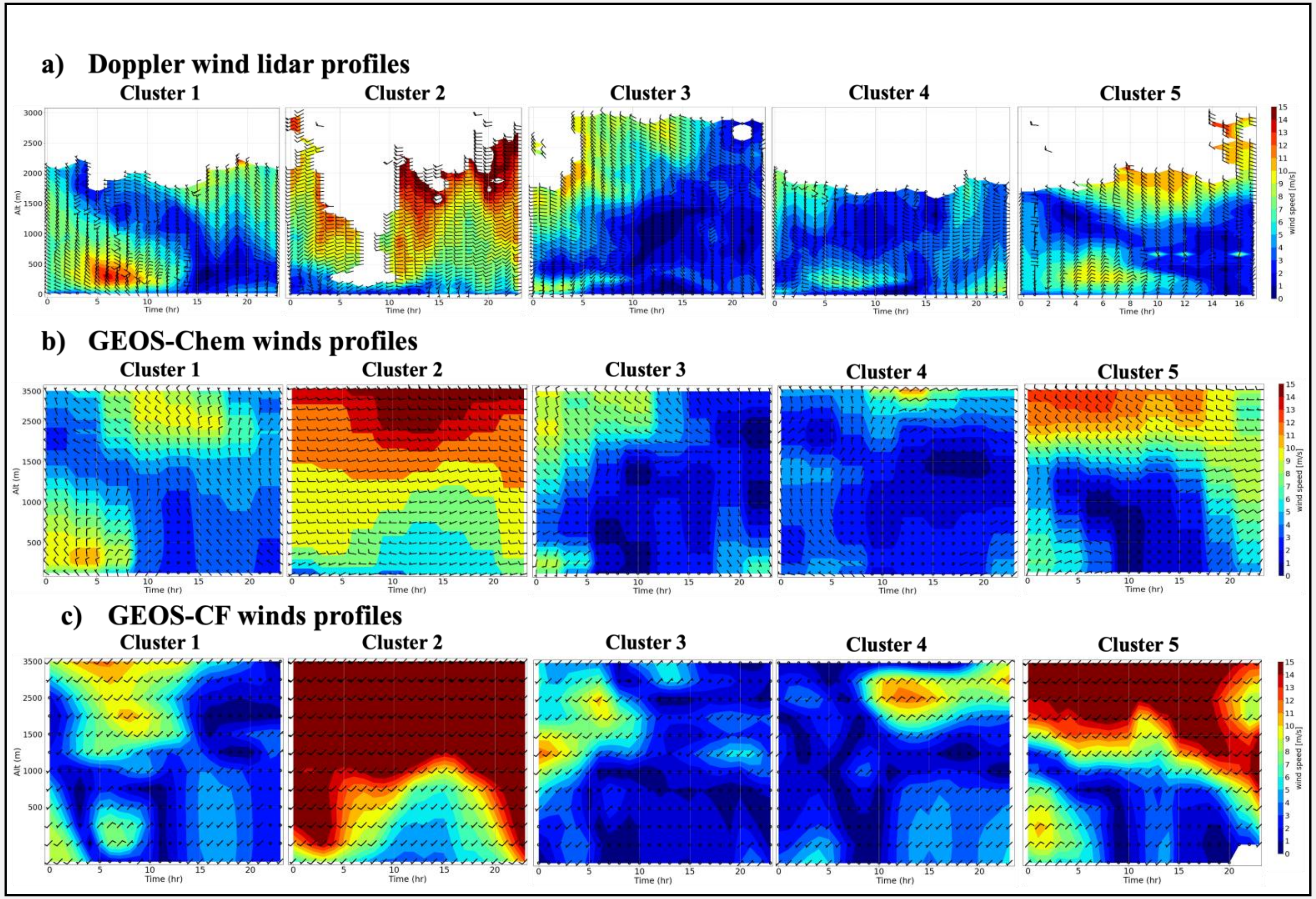
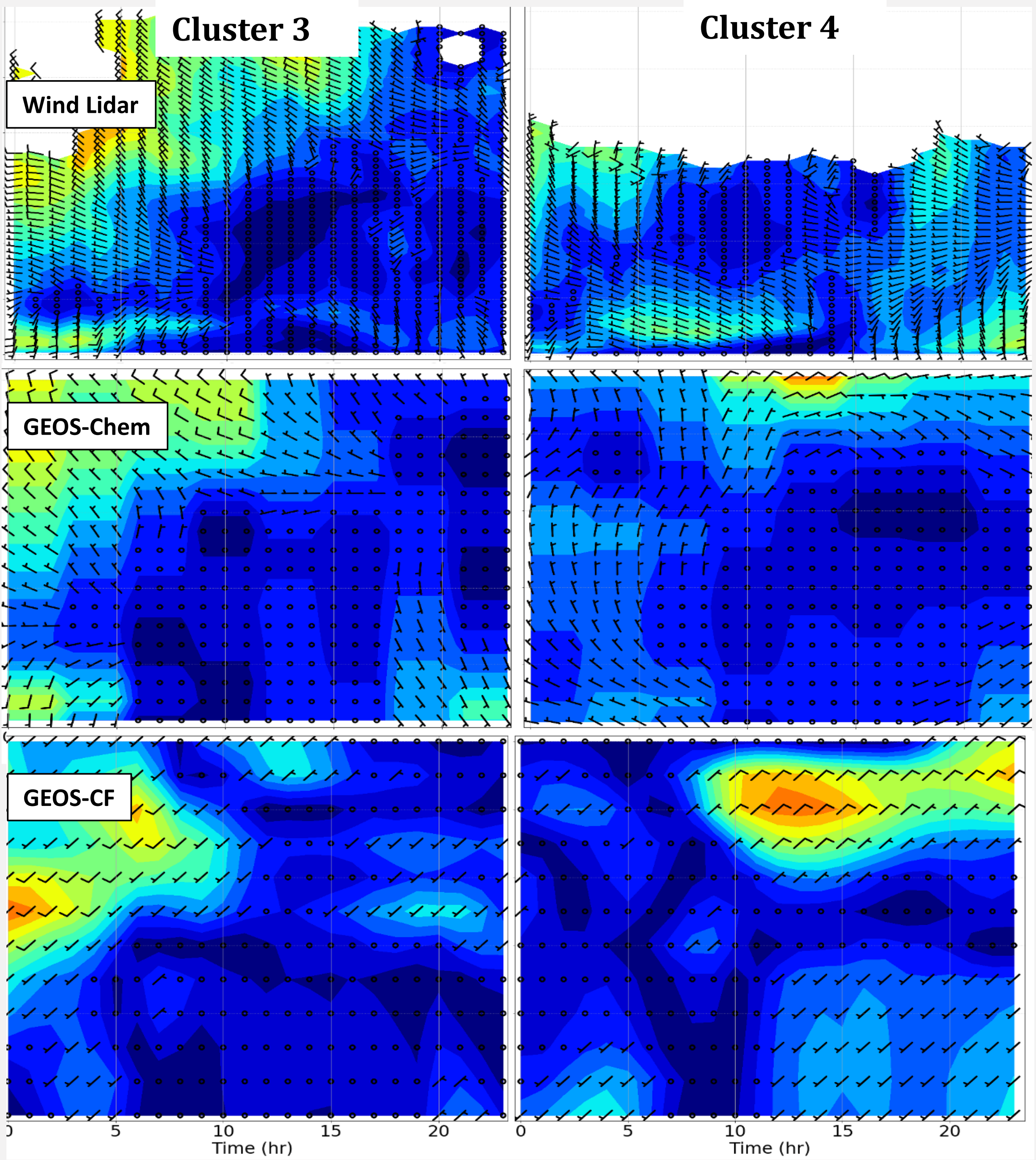


Fig. Vertical profiles of wind speed and direction from a) Doppler wind lidar, b) GEOS-Chem, and c) GEOS-CF model results from OWLETS-2 at HMI. Each figure depicts a date assigned to a specific cluster (1 - 5).

- Wind lidar data mostly limited (< 2000 m)
- Evaluated Doppler wind lidar vs. simulated wind profiles
- General under prediction in windspeeds
- Poorest model performance:
  - Cluster 3 & 4
- Both models had the highest bias & lowest correlation simulating low-level O_3 in <u>*Cluster 3*</u>
- The models struggle to capture the finer processes captured by lidar



Sea Breeze cases

- Midday wind deceleration & directional shift (offshore → onshore) at 11:00 LT accompanied by calm winds (0 m/s)
- Indicating possible common sea/bay breeze event
- Sea breeze events give cause to enhanced surface O_3
- Timing of event poorly simulated
- Models predict little/no wind much earlier in profile (begin at 06:00 LT) & no well-defined shift
- Coarser model resolution = not able to capture phenomena in these cases
- Could have facilitated in high O_3 biases for these clusters

- Although GEOS-CF has a finer horizontal resolution than GEOS-Chem, results do not reveal advantages simulating wind speed/direction
- Affirms GEOS-CF spatial resolution (~25 km) still not fine enough for processes such as the sea/bay breeze

Conclusion

- Clustering approach allowed us to characterize a range of variable vertical & temporal coastal O_3 behavior for the duration of these campaigns
- Can be a good indicator of how O_3 behaves in general in coastal regions during summer months
- Clustering analysis provided an abridged method to evaluate performance of two CTMs, GEOS-Chem & GEOS-CF
- Curated clusters reveal current limitations & cases which CTMs fare simulate coastal O_3 well
- Model simulations struggle to reproduce the O_3 vertical profile magnitude, specifically in the mid-level
- Fine processes (sea/bay breeze) were still difficult to simulate even with grid resolutions …

References:

- A., Little Roderick J, and Donald B. Rubin. *Statistical Analysis with Missing Data*. Second ed., Wiley, 1987.
- Alonso, A.M., et al. “Time Series Clustering Based on Forecast Densities.” *Computational Statistics & Data Analysis*, vol. 51, no. 2, 2006, pp. 762–776., https://doi.org/10.1016/j.csda.2006.04.035.
- Banta, R. M., Senff, C. J., Nielsen-Gammon, J., Darby, L. S., Ryerson, T. B., Alvarez, R. J., Sandberg, S. P., Williams, E. J., & Trainer, M. (2005). A bad air day in Houston. *Bulletin of the American Meteorological Society*, 86(5), 657–670. https://doi.org/10.1175/BAMS-86-5-657
- Bernier, Claudia, et al. “Clustering Surface Ozone Diurnal Cycles to Understand the Impact of Circulation Patterns in Houston, TX.” *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 23, 2019, pp. 13457–13474., https://doi.org/10.1029/2019jd031725.
- Caicedo, Vanessa., et al. “Bay Breeze and Sea Breeze Circulation Impacts on the Planetary Boundary Layer and Air Quality from an Observed and Modeled Discover-Aq Texas Case Study.” *Journal of Geophysical Research: Atmospheres*, 2019, https://doi.org/10.1029/2019jd030523.
- Charrad, Malika, et al. “NBCLUST: AnRpackage for Determining the Relevant Number of Clusters in a Data Set.” *Journal of Statistical Software*, vol. 61, no. 6, 2014, https://doi.org/10.18637/jss.v061.i06.
- Christiansen, Bo. “Atmospheric Circulation Regimes: Can Cluster Analysis Provide the Number?” *Journal of Climate*, vol. 20, no. 10, 2007, pp. 2229–2250., https://doi.org/10.1175/jcli4107.1.
- Coggon, Matthew M; Gkatzelis, Georgios I; McDonald, Brian C; Gilman, Jessica B; Schwantes, Rebecca H; Abuhassan, Nader; Aikin, Kenneth C; Arend, Mark F; Berkoff, Timothy A; Brown, Steven S; “Volatile chemical product emissions enhance ozone and modulate urban chemistry”, *Proceedings of the National Academy of Sciences*, 118,32,2021, National Academy of Sciences
- Couillard, Maxim H; Schwab, Margaret J; Schwab, James J; Lu, Cheng-Hsuan; Joseph, Everette; Stutsrim, Brennan; Shrestha, Bhupal; Zhang, Jie; Knepp, Travis N; Gronoff, Guillaume P; “Vertical Profiles of Ozone Concentrations in the Lower Troposphere Downwind of New York City during LISTOS 2018-2019”, *Journal of Geophysical Research: Atmospheres*, e2021JD035108,2021,
- Darby, Lisa S. “Cluster Analysis of Surface Winds in Houston, Texas, and the Impact of Wind Patterns on Ozone.” *Journal of Applied Meteorology*, vol. 44, no. 12, 2005, pp. 1788–1806., https://doi.org/10.1175/jam2320.1.
- Davis, Robert E., et al. “A Comparison of Trajectory and Air Mass Approaches to Examine Ozone Variability.” *Atmospheric Environment*, vol. 44, no. 1, 2010, pp. 64–74., https://doi.org/10.1016/j.atmosenv.2009.09.038.
- De Young, Russell, et al. “Langley Mobile Ozone LIDAR: Ozone and Aerosol Atmospheric Profiling for Air Quality Research.” *Applied Optics*, vol. 56, no. 3, 2017, p. 721., https://doi.org/10.1364/ao.56.000721.
- Dreessen, Joel, et al. “Observed Ozone over the Chesapeake Bay Land-Water Interface: The Hart-Miller Island Pilot Project.” *Journal of the Air & Waste Management Association*, vol. 69, no. 11, 2019, pp. 1312–1330., https://doi.org/10.1080/10962247.2019.1668497.
- EPA NEI (National Emissions Inventory v1): Air Pollutant Emission Trends Data, available at: http://www.epa.gov/ttn/chief/trends/index.html last access: 23 June 2015.
- Farris, Betsy M., et al. “Demonstration of an off-Axis Parabolic Receiver for near-Range Retrieval of Lidar Ozone Profiles” *Atmospheric Measurement Techniques*, vol. 12, no. 1, 2019, pp. 363–370., https://doi.org/10.5194/amt-12-363-2019.
- Gelaro, Ronald, et al. “The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (Merra-2).” *Journal of Climate*, vol. 30, no. 14, 2017, pp. 5419–5454., https://doi.org/10.1175/jcli-d-16-0758.1.
- Gronoff, Guillaume, et al. “A Method for Quantifying near Range Point Source Induced O3 Titration Events Using Co-Located Lidar and Pandora Measurements.” *Atmospheric Environment*, vol. 204, 2019, pp. 43–52., https://doi.org/10.1016/j.atmosenv.2019.01.052.
- Gronoff, G; Berkoff, T; Knowland, KE; Lei, L; Shook, M; Fabbri, B; Carrion, W; Langford, AO; “Case study of stratospheric Intrusion above Hampton, Virginia: lidar-observation and modeling analysis”, *Atmospheric Environment*,118498,2021,Pergamon
- Han, Jiawei, et al. *Data Mining Concepts and Techniques*. Morgan Kaufmann, 2011.
- Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., et al. (2018). Global simulation of tropospheric chemistry at 12.5 km resolution: Performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM). *Geoscientific Model Development*, 11, 4603-4620. https://doi.org/10.5194/gmd-11-4603-2018.

- Kaufman, L. & Rousseeuw, P. (1990). Finding groups in data: An introduction to cluster analysis.
- Keller, Christoph A., et al. “Description of the NASA Geos Composition Forecast Modeling System GEOS-CF v1.0.” *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 4, 2021, https://doi.org/10.1029/2020ms002413.
- Knowland, K.E., Keller, C.A., Lucchesi, R., 2019. ‘File specification for GEOS-CF products’, GMAO office note No. 17 (version 1.0). available from: https://gmao.gsfc. nasa.gov/pubs/office_notes.php, 32.
- Knowland, K. Emma, et al. “NASA GEOS Composition Forecast Modeling System GEOS-CF v1.0: Stratospheric Composition.” 2021, https://doi.org/10.1002/essoar.10508148.1.
- Lawson, Richard G., and Peter C. Jurs. “New Index for Clustering Tendency and Its Application to Chemical Problems.” *Journal of Chemical Information and Computer Sciences*, vol. 30, no. 1, 1990, pp. 36–41., https://doi.org/10.1021/ci00065a010.
- Leblanc, Thierry, et al. "Validation of the TOLNet lidars: the Southern California Ozone Observation Project (SCOOP)." *Atmospheric measurement techniques* 11.11 (2018): 6137-6162.
- Lei, Liqiao; Berkoff, Timothy A; Gronoff, Guillaume P; Su, Jia; Nehrir, Amin R; Wu, Yonghua; Moshary, Fred; Kuang, Shi; “Retrieval of UVB aerosol extinction profiles from the ground-based Langley Mobile Ozone Lidar (LMOL) system”, *Atmospheric Measurement Techniques Discussions*, 1-22,2021, Copernicus GmbH
- Li, Wei, et al. “Identification of Sea Breeze Recirculation and Its Effects on Ozone in Houston, TX, during Discover-Aq 2013.” *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 22, 2020, https://doi.org/10.1029/2020jd033165.
- McDuffie, Erin E., et al. “A Global Anthropogenic Emission Inventory of Atmospheric Pollutants from Sector- and Fuel-Specific Sources (1970–2017): An Application of the Community Emissions Data System (CEDS).” *Earth System Science Data*, vol. 12, no. 4, 2020, pp. 3413–3442., https://doi.org/10.5194/essd-12-3413-2020.
- Orbe, Clara, et al. “Large-Scale Atmospheric Transport in GEOS Replay Simulations.” *Journal of Advances in Modeling Earth Systems*, vol. 9, no. 7, 2017, pp. 2545–2560., https://doi.org/10.1002/2017ms001053.
- Ring, Allison M., et al. “Evaluating Commercial Marine Emissions and Their Role in Air Quality Policy Using Observations and the CMAQ Model.” *Atmospheric Environment*, vol. 173, 2018, pp. 96–107., https://doi.org/10.1016/j.atmosenv.2017.10.037.
- Stauffer, Ryan M., et al. “Characterizing Global Ozonesonde Profile Variability from Surface to the UT/Ls with a Clustering Technique and MERRA-2 Reanalysis.” *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 11, 2018, pp. 6213–6229., https://doi.org/10.1029/2018jd028465.
- Strode, Sarah A., et al. “Global Changes in the Diurnal Cycle of Surface Ozone.” *Atmospheric Environment*, vol. 199, 2019, pp. 323–333., https://doi.org/10.1016/j.atmosenv.2018.11.028.
- Sullivan, J. T., et al. “Optimization of the GSFC Tropoz Dial Retrieval Using Synthetic Lidar Returns and Ozonesondes – Part 1: Algorithm Validation.” *Atmospheric Measurement Techniques*, vol. 8, no. 10, 2015, pp. 4133–4143., https://doi.org/10.5194/amt-8-4133-2015.
- Sullivan, John T., et al. “Results from the NASA GSFC and LARC Ozone Lidar Intercomparison: New Mobile Tools for Atmospheric Research.” *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 10, 2015, pp. 1779–1795., https://doi.org/10.1175/jtech-d-14-00193.1.
- Thompson, Anne M., et al. “Ozone Profiles in the Baltimore-Washington Region (2006–2011): Satellite Comparisons and Discover-AQ Observations.” *Journal of Atmospheric Chemistry*, vol. 72, no. 3-4, 2014, pp. 393–422., https://doi.org/10.1007/s10874-014-9283-z.
- Torgo, Luís. “Data Mining with R.” *Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, 2010, https://doi.org/10.1201/b10328.
- Wang, Lihua, et al. “Quantifying Tolnet Ozone Lidar Accuracy during the 2014 Discover-AQ and Frappé Campaigns.” *Atmospheric Measurement Techniques*, vol. 10, no. 10, 2017, pp. 3865–3876., https://doi.org/10.5194/amt-10-3865-2017.
- Wu, Yonghua; Nehrir, Amin R; Ren, Xinrong; Dickerson, Russell R; Huang, Jianping; Stratton, Phillip R; Gronoff, Guillaume; Kooi, Susan A; Collins, James E; Berkoff, Timothy A; "Synergistic aircraft and ground observations of transported wildfire smoke and its impact on air quality in New York City during the summer 2018 LISTOS campaign”, *Science of The Total Environment*, 773,,145030,2021,Elsevier

Extra:

Clustering application

- Features evaluated for cluster tendency (to confirm dataset contained meaningful clusters)
- Hopkins statistic which measures whether there is uniform distribution (spatial randomness) within dataset (Lawson & Jurs, 1990)
- Results: value higher than 0.75 (actual = 0.77) which by this standard indicates a clustering tendency at the 90% confidence level
- Evaluating different feature options did not lead to better statistical results than with the final chosen features
- *Nbclust* (Charrad et al., 2014) used: applies 30 indices for determining optimal number of clusters
- Tested quality of clustering results using silhouette method (Kaufman & Rousseeuw, 1990)
- Results: 6 clusters – later 5
- K-Means clustering algorithm based on Euclidean distance to each centroid
  - input data normalized (to a mean of zero & standard deviation of one)
  - to ensure each feature is given same importance in clustering (Aksoy & Haralick, 2001; Larose, 2005)

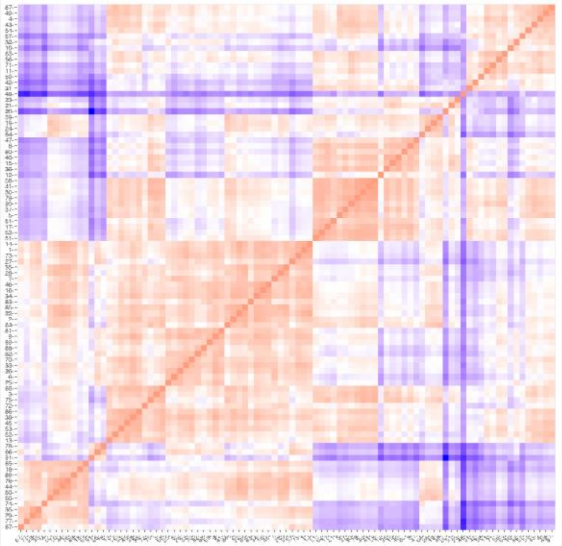


Figure S1. Visual assessment of cluster tendency (VAT) approach. Dataset high similarity (red) and low similarity (blue).

Missing data

- Input features were tailored based on structure of lidar measurements (limitations), remaining data still had missing data points
- Most cases of missing data points:
  - earlier morning measurements (06:00 – 12:00 EDT)
  - later evening measurements (16:00 – 21:00 EDT)
- 51/91 O_3 profile curtains had at least <u>*one*</u> missing data point (feature)
- Complete case analysis (CCA) (e.g. only 40 O_3 curtains)
- Single imputation (SI) technique (*knnImputation)* to fill in missing data
- Allows use of full 91 profile curtains of O_3 data ~ imputated
- Silhouette method used to test quality of newly imputed dataset
- Proved to be no worse, nor better, than CCA (*real data*) results

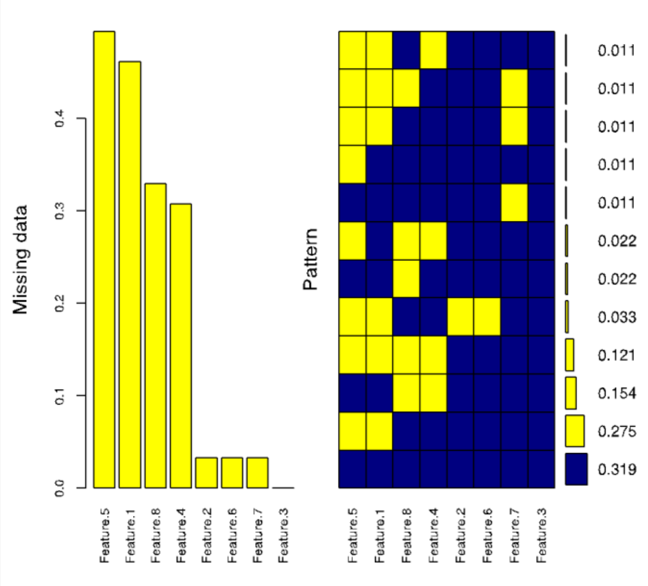


Figure S2. Percentage and pattern of missing data points by each feature used for clustering.